10 POWER CALCULATIONS FOR THE STATISTICAL TESTS

10.1 Statistical Power and the Probability of Survey Unit Release

The concept of the statistical power of a test was introduced in Section 2.3.2. The use of this concept in optimizing the design of final status surveys was discussed in Section 3.8.1. The power of a statistical test is defined as the probability that the null hypothesis is rejected when it is false. It is $1-\beta$, where β is the Type II error of the test.

The statistical power will have different implications for survey unit release, depending on whether Scenario A or B is used. The same information can be expressed slightly differently. In this report, it is expressed as the probability that the survey unit passes the statistical test, i.e., the result of the test is the decision that the survey unit may be released.

The relationship between this probability and the Type I and Type II errors was given in Table 3.1. Figures 3.9 through 3.12 show this probability as a function of the true residual radioactivity concentration for selected values of α and β over a range of sample sizes. In many cases, it will be sufficient to check the curve in these figures that corresponds most closely to the situation at hand. In the following sections, the assumptions made and the calculations performed in creating these figures are described.

10.2 Power of the Sign Test Under Scenario A

Recall that for the Sign test in Scenario A, the test statistic, S+, was equal to the number of survey unit measurements below the $DCGL_W$. If S+ exceeds the critical value *k*, then the null hypothesis that the median concentration in the survey unit exceeds the $DCGL_W$ is rejected, i.e., the survey unit passes this test. The probability that any single survey unit measurement falls below the $DCGL_W$ is found from Equation 9-2 or 9-3. The probability that more than *k* of the *N* survey unit measurements fall below the $DCGL_W$ is simply the following binomial probability:

$$\sum_{i=k+1}^{N} \binom{N}{i} [p]^i [1-p]^{N-i} = 1 - \sum_{i=0}^{k} \binom{N}{i} [p]^i [1-p]^{N-i} \approx 1 - \Phi\left(\frac{k-Np}{\sqrt{Np(1-p)}} \right) \qquad (10\text{-}1)$$

The indicated approximation is generally used when both Np and $N(1-p)$ are five or greater. $\Phi(z)$ is the cumulative standard normal distribution function given in Table A.1.

With *p* calculated as in Section 9.2, Equation 10-1 yields the probability that the null hypothesis is rejected when the true median of the residual radioactivity concentration in the survey unit is at the LBGR. This is the power of the test at the LBGR.

The probability, $p(C)$, that any single survey unit measurement falls below the $DCGL_W$ when the survey unit median concentration is at any other value, *C*, can be determined by simply replacing the value of the *LBGR* in Equation 9-2 with the value of *C*:

$$p(C) = \frac{1}{\sqrt{2\pi}\ \sigma} \int_{-\infty}^{DCGL_W} e^{-(x-C)^2/2\sigma^2} dx$$

$$= \frac{1}{\sqrt{2\pi}\ \sigma} \int_{-\infty}^{C+(DCGL_W - C)} e^{-(x-C)^2/2\sigma^2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{(DCGL_W - C)}{\sigma}} e^{-x^2/2} dx$$

$$= \Phi\left(\frac{(DCGL_W - C)}{\sigma}\right) \qquad (10\text{-}2)$$

Note that if $C = DCGL_W$, $p(C) = 0.5$. The assumption of normality is not critical in the above calculations, since it is only being used to estimate the power. However, if a different distribution is considered more appropriate, Equation 9-3 can be used to calculate *p(C)*.

When the value of *p(C)* from Equation 10-2 is inserted in Equation 10-1, we obtain the probability that the null hypothesis is rejected at the concentration *C*. When $C = DCGL_W$, this probability is the probability of a Type I error, α[1]. This calculation can even be performed for values of *C* greater than the $DCGL_W$. The probability obtained is still the probability that the null hypothesis is rejected, i.e., that the survey unit passes the test.

If the probability that the null hypothesis is rejected (calculated from Equation 10-1) is plotted against the concentration, C, the result is called a power curve. When the power calculation is performed at the design stage, using an estimated value of σ, it is called a prospective power curve. When the calculation is performed after the survey, using the standard deviation of the survey unit measurements as an estimate of σ, it is called a retrospective power curve.

To illustrate the construction of a power curve, consider the example of Chapter 5. The $DCGL_W$ for this example was 15.9 and the LBGR was 11.5. The DQOs for $\alpha = \beta = 0.05$ resulted in a sample size of $N = 21$, using the estimate that $\sigma = 3.3$. From Table A.3, the critical value for the Sign test with N=21 and $\alpha = 0.05$ is $k = 14$. This is all of the information necessary to construct the prospective power curve. To construct the retrospective power curve, we use the standard deviation of the measurement data, 9.5, as the estimate of σ.

The results of these calculations are shown in Table 10.1 and Figure 10.1.

---

[1]The value of α actually obtained from Equation 10-1 should be close to that specified in the DQOs. It may not exactly equal that value when the sample sizes are small, since the critical value, k, can only take integer values.

Table 10.1 Example Power Calculations: Sign Test Scenario A

	Prospective			Retrospective		
C	**($DCGL_W$ − *C*)/σ**	***p*(*C*))** (Eq. 10-2)	**power** (Eq. 10-1)	**($DCGL_W$ − C)/σ**	***p*(*C*)** (Eq. 10-2)	**power** (Eq. 10-1)
0	4.82	1.0000	1.000	1.67	0.9525	1.000
5	3.30	0.9995	1.000	1.15	0.8749	0.989
6	3.00	0.9987	1.000	1.04	0.8508	0.972
7	2.70	0.9965	1.000	0.94	0.8264	0.942
8	2.39	0.9916	1.000	0.83	0.7967	0.884
9	2.09	0.9817	1.000	0.73	0.7673	0.802
10	1.79	0.9633	1.000	0.62	0.7324	0.679
11	1.48	0.9306	1.000	0.52	0.6985	0.544
11.5	1.33	0.9082	0.998	0.46	0.6772	0.459
12	1.18	0.8810	0.991	0.41	0.6591	0.390
13	0.88	0.8106	0.914	0.31	0.6217	0.262
14	0.58	0.7190	0.627	0.20	0.5793	0.151
15	0.27	0.6064	0.217	0.09	0.5359	0.076
15.9	0.00	0.5000	0.039	0.00	0.5000	0.039
16	−0.03	0.4880	0.031	−0.01	0.4960	0.036
17	−0.33	0.3707	0.001	−0.12	0.4522	0.014
18	−0.64	0.2611	0.000	−0.22	0.4129	0.005
19	−0.94	0.1736	0.000	−0.33	0.3707	0.001
20	−1.24	0.1075	0.000	−0.43	0.3336	0.000

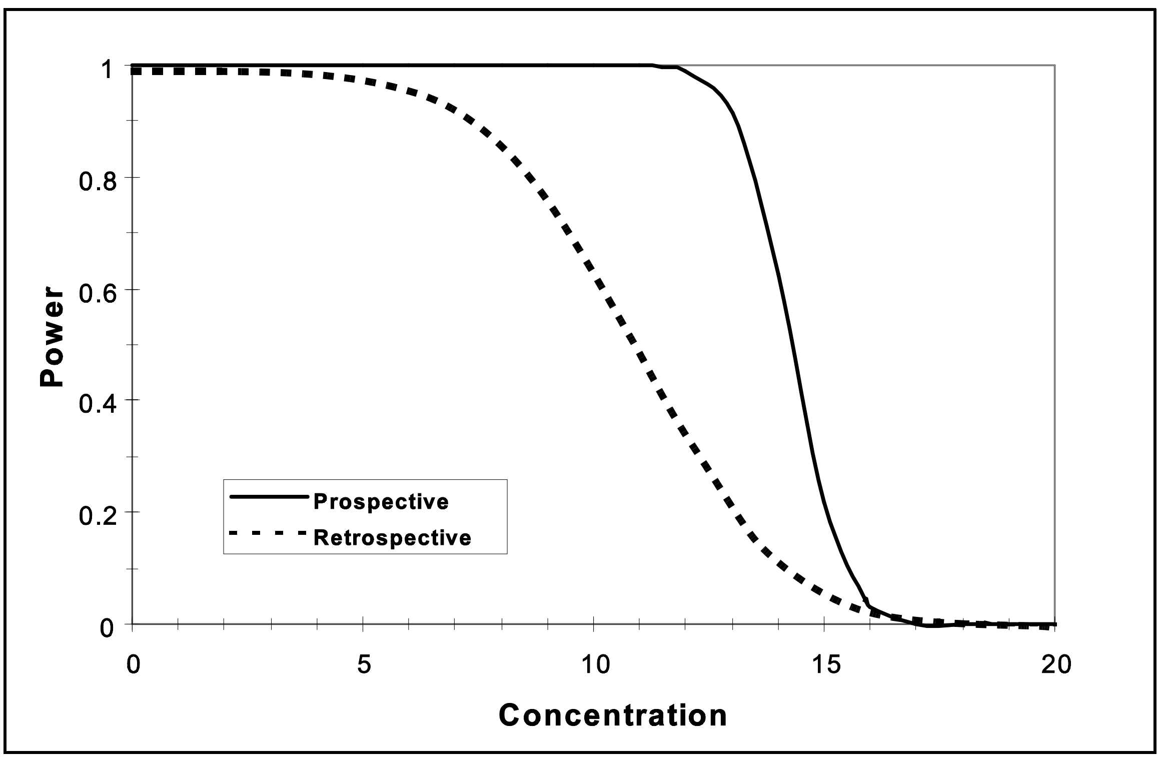

Figure 10.1 Example Power Curves: Sign Test Scenario A

Notice that the increase of σ due to a higher than anticipated measurement standard deviation causes the retrospective power curve to differ considerably from the prospective power curve. In Table 3.3, we see that $\Delta/\sigma = (15.9 - 11.5)/9.5 = 0.46$ results in a much larger required sample size (over 100) to achieve the desired power. Recall that in this example, S+ = 11, which is smaller than the critical value $k = 14$. Thus the null hypothesis was not rejected. The survey unit did not pass. We now see that this might have been a consequence of having insufficient power rather than the survey unit actually exceeding the release criterion. The lack of power was due to underestimating the measurement variability.

10.3 Power of the Sign Test Under Scenario B

Recall that for the Sign test in Scenario B, the test statistic, S+, was equal to the number of survey unit measurements above the LBGR. If S+ exceeds the critical value *k*, then the null hypothesis that the median concentration in the survey unit is less than the LBGR is rejected, i.e. the survey unit does not pass. The probability that any single survey unit measurement falls below the $DCGL_W$, is found from Equation 9-4 or 9-5. The probability that more than *k* of the *N* survey unit measurements fall above the LBGR is simply the following binomial probability:

$$\sum_{i=k+1}^{N} \binom{N}{i} [p]^i [1-p]^{N-i} = 1 - \sum_{i=0}^{k} \binom{N}{i} [p]^i [1-p]^{N-i} \approx 1 - \Phi\left(\frac{k - Np}{\sqrt{Np(1-p)}} \right) \qquad (10\text{-}3)$$

The indicated approximation is generally used when both *Np* and $N(1-p)$ are five or greater. $\Phi(z)$ is the cumulative standard normal distribution function given in Table A.1.

With *p* calculated as in Section 9.3, this is the probability that the null hypothesis is rejected when the true median of the residual radioactivity concentration in the survey unit is at the $DCGL_W$. This is the power of the test at the $DCGL_W$.

The probability, *p(C)*, that any single survey unit measurement falls above the LBGR when the survey unit median concentration is at any other value, *C*, can be determined by simply replacing the value of the $DCGL_W$ in Equation 9-4 with the value of *C*:

$$p(C) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{LBGR}^{\infty} e^{-(x-C)^2/2\sigma^2} dx$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{(LBGR-C)+C}^{-\infty} e^{-(x-C)^2/2\sigma^2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{LBGR-C}{\sigma}}^{\infty} e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{C - LBGR}{\sigma}} e^{-x^2/2} dx$$

$$= \Phi\left(\frac{C - LBGR}{\sigma}\right) \tag{10-4}$$

Note that if $C = LBGR$, $p(C) = \Phi(0) = 0.5$. The assumption of normality is not critical in the preceding calculations, since it is only being used to estimate the power. However, if a different distribution is considered more appropriate, Equation 9-5 can be used to calculate *p(C)*.

When the value of *p(C)* from Equation 10-4 is inserted into Equation 10-2, we obtain the probability that the null hypothesis is rejected at the concentration, *C*. When *C = LBGR*, this probability is the probability of a Type I error, α[2]. This calculation can even be performed for values of *C* less than the LBGR. The probability obtained is still the probability that the null hypothesis is rejected, i.e., that the survey unit passes the test, but it is not normally referred to as the power.

If the probability that the null hypothesis is rejected (calculated from Equation 10-3) is plotted against the concentration, *C*, the result is called a power curve. When the power calculation is performed at the design stage, using an estimated value of σ, it is called a prospective power curve. When the calculation is performed after the survey, using the standard deviation of the survey unit measurements as an estimate of σ, it is called a retrospective power curve.

To illustrate the construction of a power curve, consider the example of Chapter 5. The $DCGL_W$ for this example was 15.9 and the LBGR was 11.5. The DQOs for $\alpha = \beta = 0.05$ resulted in a sample size of $N = 21$, using the estimate that $\sigma = 3.3$. From Table A.3, the critical value for the Sign test with $N = 21$ and $\alpha = 0.05$ is $k = 14$. This is all of the information necessary to construct the prospective power curve. To construct the retrospective power curve, we use the standard deviation of the measurement data, 9.5, as the estimate of σ. The results of these calculations are shown in Table 10.2 and Figure 10.2.

Notice that the increase of σ due to a higher than anticipated measurement standard deviation causes the retrospective power curve to differ considerably from the prospective power curve. $\Delta/\sigma = (15.9 - 11.5)/9.5 = 0.46$ results in a much larger required sample size to achieve the desired power. Recall that in this example, S+ = 13, which is smaller than the critical value, $k = 14$. Thus the null hypothesis was not rejected. The survey unit passes. We now see that this might have been a consequence of having insufficient power rather than the survey unit actually meeting the release criterion. The lack of power was due to underestimating the measurement variability.

---

[2] The value of α actually obtained from Equation 10-2 should be close to that specified in the DQOs. It may not exactly equal that value when the sample sizes are small, since the critical value, *k*, can only take integer values.

Table 10.2 Example Power Calculations: Sign Test Scenario B

	Prospective			Retrospective		
C	**(*C* − *LBGR*)/σ**	***p*(*C*)** (Eq. 10-3)	**power** (Eq. 10-4)	**(*C* − *LBGR*)/σ**	***p*(*C*)** (Eq. 10-2)	**power** (Eq. 10-1)
0	−3.48	0.0003	0.000	−1.21	0.1131	0.000
5	−1.97	0.0244	0.000	−0.68	0.2483	0.000
6	−1.67	0.0475	0.000	−0.58	0.2810	0.000
7	−1.36	0.0869	0.000	−0.47	0.3192	0.000
8	−1.06	0.1446	0.000	−0.37	0.3557	0.001
9	−0.76	0.2236	0.000	−0.26	0.3974	0.003
10	−0.45	0.3264	0.000	−0.16	0.4364	0.009
11	−0.15	0.4404	0.010	−0.05	0.4801	0.026
11.5	0.00	0.5000	0.039	0.00	0.5000	0.039
12	0.15	0.5596	0.112	0.05	0.5199	0.057
13	0.45	0.6736	0.445	0.16	0.5636	0.119
14	0.76	0.7764	0.830	0.26	0.6026	0.207
15	1.06	0.8554	0.976	0.37	0.6443	0.336
15.9	1.33	0.9082	0.998	0.46	0.6772	0.459
16	1.36	0.9131	0.999	0.47	0.6808	0.474
17	1.67	0.9525	1.000	0.58	0.7190	0.627
18	1.97	0.9756	1.000	0.68	0.7517	0.750
19	2.27	0.9884	1.000	0.79	0.7852	0.854
20	2.58	0.9951	1.000	0.89	0.8133	0.919
21	2.88	0.9980	1.000	1.00	0.8413	0.962
22	3.18	0.9993	1.000	1.11	0.8665	0.984
23	3.48	0.9997	1.000	1.21	0.8869	0.993
24	3.79	0.9999	1.000	1.32	0.9066	0.998
25	4.09	1.0000	1.000	1.42	0.9222	0.999

In Scenario A, the power and the probability that the survey unit passes the test are equivalent. In Scenario B, the power is equivalent to the probability that the survey unit does not pass. To plot the probability that the survey unit passes, the power is subtracted from 1. The result is shown in Figure 10.3.

10.4 Power of the Wilcoxon Rank Sum Test Under Scenario A

Recall that for the Wilcoxon Rank Sum (WRS) test in Scenario A, the test statistic, W_r, was equal to the sum of the ranks of the reference area measurements adjusted for the $\mathrm{DCGL_W}$. If W_r exceeds the critical value W_c, then the null hypothesis that the median concentration in the survey unit exceeds that in the reference area by more than the $\mathrm{DCGL_W}$ is rejected, i.e., the survey unit passes this test.

The power of the WRS test is very difficult to calculate exactly. However, a good approximation is available (Lehmann and D'Abrera, 1975,***p*(*C*)** Chapter 2, Section 3, pp. 69–75).

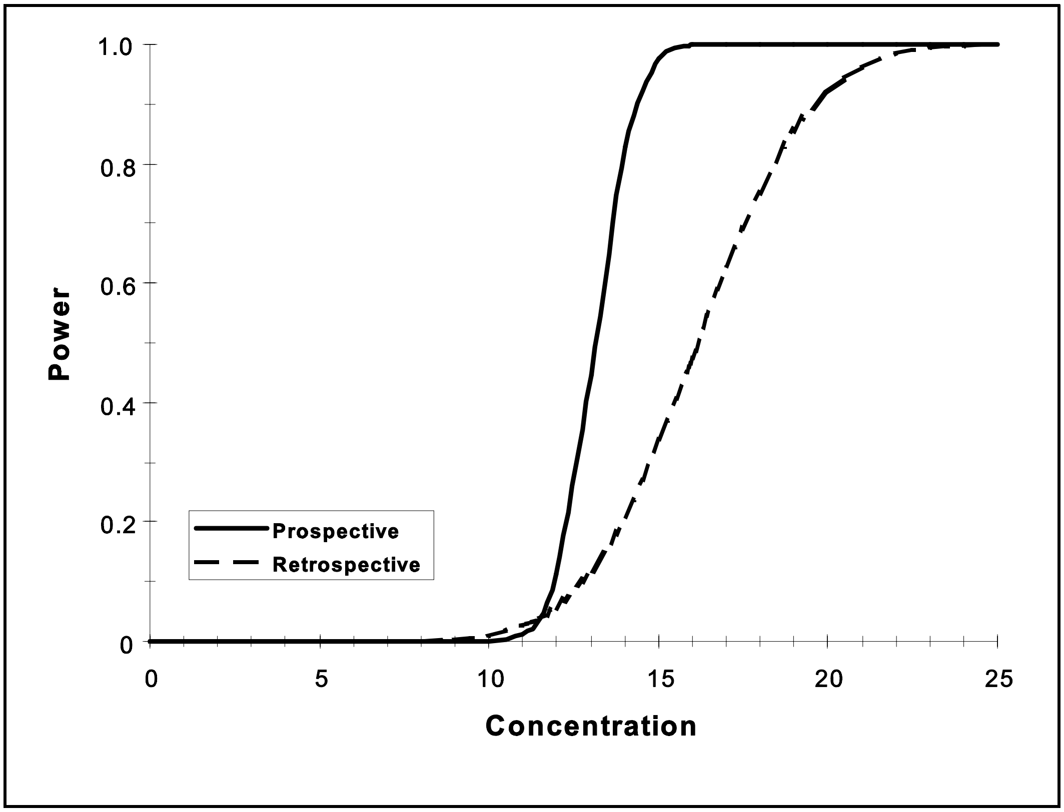

Figure 10.2 Example Power Curves: Sign Test Scenario B

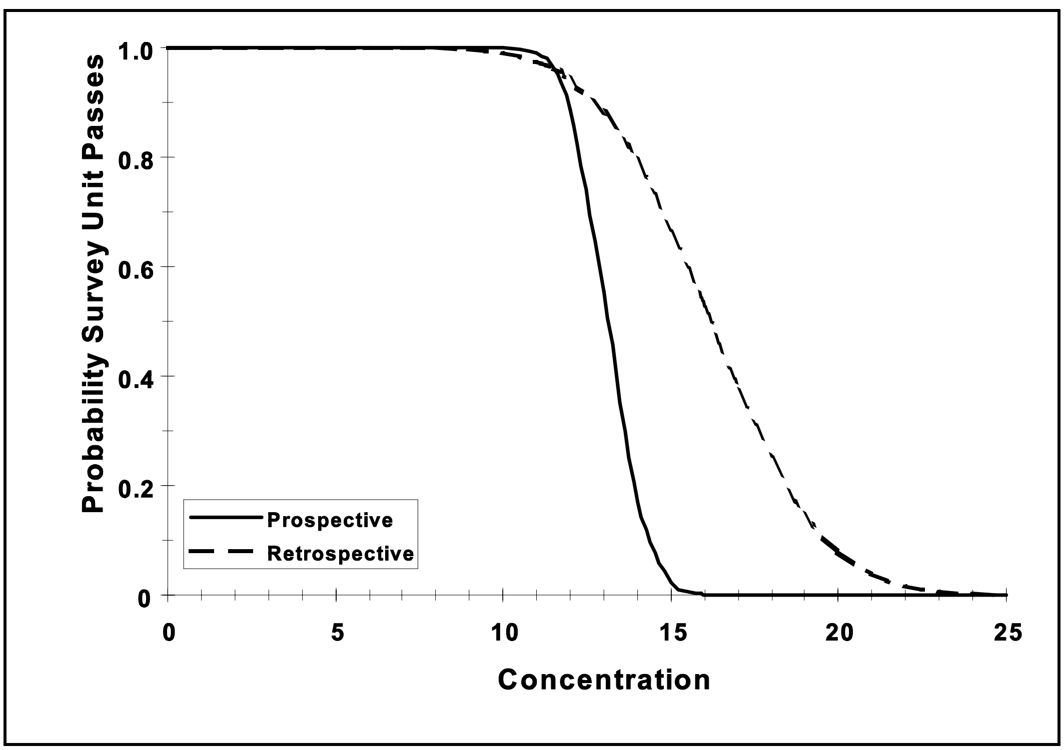

Figure 10.3 Probability Example Survey Unit Passes: Sign Test Scenario B

If the distribution of the Mann-Whitney form of the WRS test statistic is approximated by a normal distribution, the probability that the null hypothesis will be rejected when the alternative is true can be calculated from:

$$\text{Power} = 1 - \Phi\left[\frac{W_c - 0.5 - 0.5m(m+1) - \mathrm{E}(W_{MW})}{\sqrt{\mathrm{Var}(W_{MW})}}\right] \tag{10-5}$$

where W_c is the critical value found in Table A.4 for the appropriate values of the Type I error, α, the number of survey unit measurements, n, and the number of reference area measurements, m. $\mathrm{E}(W_{MW})$ and $\mathrm{Var}(W_{MW})$ are the mean and variance of the Mann-Whitney form of the WRS test statistic. Values of $\Phi(z)$, the standard normal cumulative distribution function, are given in Table A.1.

The Mann-Whitney form of the WRS test statistic is $W_{MW} = W_r - 0.5m(m+1)$. It is obtained by subtracting from W_r its minimum value, $0.5m(m+1)$. The mean of W_{MW} is

$$\mathrm{E}(W_{MW}) = mnp_1 \tag{10-6}$$

where p_1 is the probability that any single measurement from the survey unit exceeds a single measurement from the reference area by less than the $\mathrm{DCGL_W}$. This probability depends on the difference in median concentration between the survey unit and the reference area. When this difference is equal to the LBGR, then p_1 is equal to P_r as calculated from Equation 9-7. For other values of the difference median concentration between the survey unit and the reference area, C, we simply replace the LBGR in Equation 9-7 with C:

$$P_r(C) = \text{Probability}(U = X - Y < DCGL)$$

$$= \int_{-\infty}^{DCGL}\left[\int_{-\infty}^{\infty} f_X(u+y) f_Y(y)\,dy\right] du$$

$$= \int_{-\infty}^{DCGL}\left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(u+y-C-BKGD)^2/2\sigma^2} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-BKGD)^2/2\sigma^2} dy\right] du$$

$$= \frac{1}{\sqrt{2\pi}\,\sqrt{2}\,\sigma} \int_{-\infty}^{DCGL} e^{-(u-C)^2/4\sigma^2} du$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{DCGL_W - C}{\sqrt{2}\sigma}} e^{-x^2/2} dx$$

$$= \Phi\left(\frac{DCGL_W - C}{\sqrt{2}\sigma} \right) \tag{10-7}$$

Note that if $C = DCGL_W$, then $p_1(C) = 0.5$. The assumption of normality is not critical in the preceding calculations, since it is only being used to estimate the power. However, if a different distribution is considered more appropriate, Equation 9-8 can be used to calculate $p_1(C)$.

The variance of W_{MW} is:

$$\mathrm{Var}(W_{MW}) = mnp_1(1-p_1) + mn(n-1)(p_2 - p_1^2) + mn(m-1)(p_3 - p_1^2) \tag{10-8}$$

p_2 is the probability that two random measurements from the survey unit will each exceed a single random measurement from the reference area by less than the $\mathrm{DCGL_W}$; and p_3 is the probability that a single random measurement from the survey unit will exceed each of two random measurements from the reference area unit by less than the $\mathrm{DCGL_W}$. When the difference in the concentration distributions of the survey unit and the reference area measurements consists of a shift in the median, and the measurement distributions are symmetric, then $p_2 = p_3$. Then the variance of W_{MW} simplifies to

$$\mathrm{Var}(W_{MW}) = mnp_1(1-p_1) + mn(n+m-2)(p_2 - p_1^2) \tag{10-9}$$

If the measurement distributions are normal, then p_2 is equal to the probability that two correlated standard normal random variables (i.e., with mean = 0 and variance = 1), with correlation coefficient 0.5, are both less than $(DCGL_W - C)/(\sigma\sqrt{2})$. This probability also depends on the difference in median concentration, C, between the survey unit and the reference area. Even with the simplifications employed, the values of p_2 are not easy to calculate. Table 10.3 provides values of p_1 and p_2 as a function of $(DCGL_W - C)/\sigma$ that can be used in calculating the mean and variance of W_{MW}. Nomographs of bivariate normal probabilities that can also be used for this purpose are given in Abramowitz and Stegun (1972).

The power calculated using Equations 10-5 through 10-8 is an approximation. This approximation was compared against the power simulations for the WRS test reported by Gilbert and Simpson (PNL-7409, 1992). It was found that the approximation is sufficiently accurate to determine if the sample design achieves the DQOs.

Table 10.3 Values of p_1 and p_2 for Computing the Mean and Variance of W_{MW} [3]

$(DCGL_W - C)/\sigma$	p_1	p_2	$(DCGL_W - C)/\sigma$	p_1	p_2
−6.0	0.000010	0.000000	0.7	0.689691	0.544073
−5.0	0.000204	0.000010	0.8	0.714196	0.574469
−4.0	0.002339	0.000174	0.9	0.737741	0.604402
−3.5	0.006664	0.000738	1.0	0.760250	0.633702
−3.0	0.016947	0.002690	1.1	0.781662	0.662216
−2.5	0.038550	0.008465	1.2	0.801928	0.689800
−2.0	0.078650	0.023066	1.3	0.821015	0.716331
−1.9	0.089555	0.027714	1.4	0.838901	0.741698
−1.8	0.101546	0.033114	1.5	0.855578	0.765812
−1.7	0.114666	0.039348	1.6	0.871050	0.788602
−1.6	0.128950	0.046501	1.7	0.885334	0.810016
−1.5	0.144422	0.054656	1.8	0.898454	0.830022
−1.4	0.161099	0.063897	1.9	0.910445	0.848605
−1.3	0.178985	0.074301	2.0	0.921350	0.865767
−1.2	0.198072	0.085944	2.1	0.931218	0.881527
−1.1	0.218338	0.098892	2.2	0.940103	0.895917
−1.0	0.239750	0.113202	2.3	0.948062	0.908982
−0.9	0.262259	0.128920	2.4	0.955157	0.920777
−0.8	0.285804	0.146077	2.5	0.961450	0.931365
−0.7	0.310309	0.164691	2.6	0.967004	0.940817
−0.6	0.335687	0.184760	2.7	0.971881	0.949208
−0.5	0.361837	0.206266	2.8	0.976143	0.956616
−0.4	0.388649	0.229172	2.9	0.979848	0.963118
−0.3	0.416002	0.253419	3.0	0.983053	0.968795
−0.2	0.443769	0.278930	3.1	0.985811	0.973725
−0.1	0.471814	0.305606	3.2	0.988174	0.977981
0.0	0.500000	0.333333	3.3	0.990188	0.981636
0.1	0.528186	0.361978	3.4	0.991895	0.984758
0.2	0.556231	0.391392	3.5	0.993336	0.987410
0.3	0.583998	0.421415	4.0	0.997661	0.995497
0.4	0.611351	0.451875	5.0	0.999796	0.999599
0.5	0.638163	0.482593	6.0	0.999989	0.999978
0.6	0.664313	0.513387			

When the values of $p_1(C)$ and $p_2(C)$ and the resulting values of $E(W_{MW})$ and $Var(W_{MW})$ are inserted in Equation 10-5, we obtain the probability that the null hypothesis is rejected at concentration C. When $C = DCGL_W$, this probability is the probability of a Type I error, α.[4].

[3]This table may also be used for Scenario B when $(DCGL_W - C)/\sigma$ is replaced by $(C - LBGR)/\sigma$.

[4] The value of α actually obtained from Equation 10-5 should be close to that specified in the DQOs. It may not exactly equal that value when the sample sizes are small, since the critical value, k, can only take integer values.

The preceding calculations can even be performed for values of C greater than the $DCGL_W$. The probability obtained is still the probability that the null hypothesis is rejected, i.e., that the survey unit passes the test.

If the probability that the null hypothesis is rejected (calculated from Equation 10-5) is plotted against the concentration, C, the result is called a power curve. When the power calculation is performed at the design stage, using an estimated value of σ, it is called a prospective power curve. When the calculation is performed after the survey, using the standard deviation of the survey unit measurements as an estimate of σ, it is called a retrospective power curve.

To illustrate the construction of a power curve, consider the example of Chapter 6. The $DCGL_W$ for this example was 160 and the LBGR was 142. The DQOs for $\alpha = \beta = 0.05$ resulted in a sample size of $n = m = 10$, using the estimate that $\sigma = 6$. Twelve samples each were actually taken from the survey unit and the reference area. From Table A.4, the critical value for the WRS test with $n = m = 12$ and $\alpha = 0.05$ is $W_c = 179$. This is all of the information necessary to construct the prospective power curve. To construct the retrospective power curve, we use the larger of the standard deviations of the measurement data from the survey unit and the reference area, 8.1, as the estimate of σ.

The results of these calculations are shown in Table 10.4 and Figure 10.4. In the figure it can be seen that the retrospective power is slightly less than that specified in the DQOs. However, in this example, the null hypothesis was rejected, so the question of the power is moot. The retrospective power calculation is really only necessary when the null hypothesis is not rejected. In that case, it is important to know that it was not rejected simply because there was insufficient power. When the null hypothesis is rejected in spite of insufficient power, the survey designer can consider himself lucky, but the conclusion is still statistically valid.

Table 10.4 Example Prospective Power Calculation: WRS Test Scenario A

C	$(DCGL_W - C)/\sigma$	p_1	p_2	$E(W_{MW})$	$Var(W_{MW})$	$SD(W_{MW})$	z	Power
136	4.00	0.997661	0.995497	143.7	0.9	0.9	−46.21	1.00
139	3.50	0.993336	0.987410	143.0	3.2	1.8	−23.96	1.00
142	3.00	0.983053	0.968795	141.6	10.0	3.2	−12.98	1.00
145	2.50	0.961450	0.931365	138.4	27.4	5.2	−7.24	1.00
148	2.00	0.921350	0.865767	132.7	63.9	8.0	−4.02	1.00
151	1.50	0.855578	0.765812	123.2	124.9	11.2	−2.03	0.98
154	1.00	0.760250	0.633702	109.5	202.8	14.2	−0.63	0.74
157	0.50	0.638163	0.482593	91.9	271.9	16.5	0.52	0.30
160	0.00	0.500000	0.333333	72.0	300.0	17.3	1.65	0.05
163	−0.50	0.361837	0.206266	52.1	271.9	16.5	2.93	0.00
166	−1.00	0.239750	0.113202	34.5	202.8	14.2	4.63	0.00
169	−1.50	0.144422	0.054656	20.8	124.9	11.2	7.13	0.00
172	−2.00	0.078650	0.023066	11.3	63.9	8.0	11.15	0.00

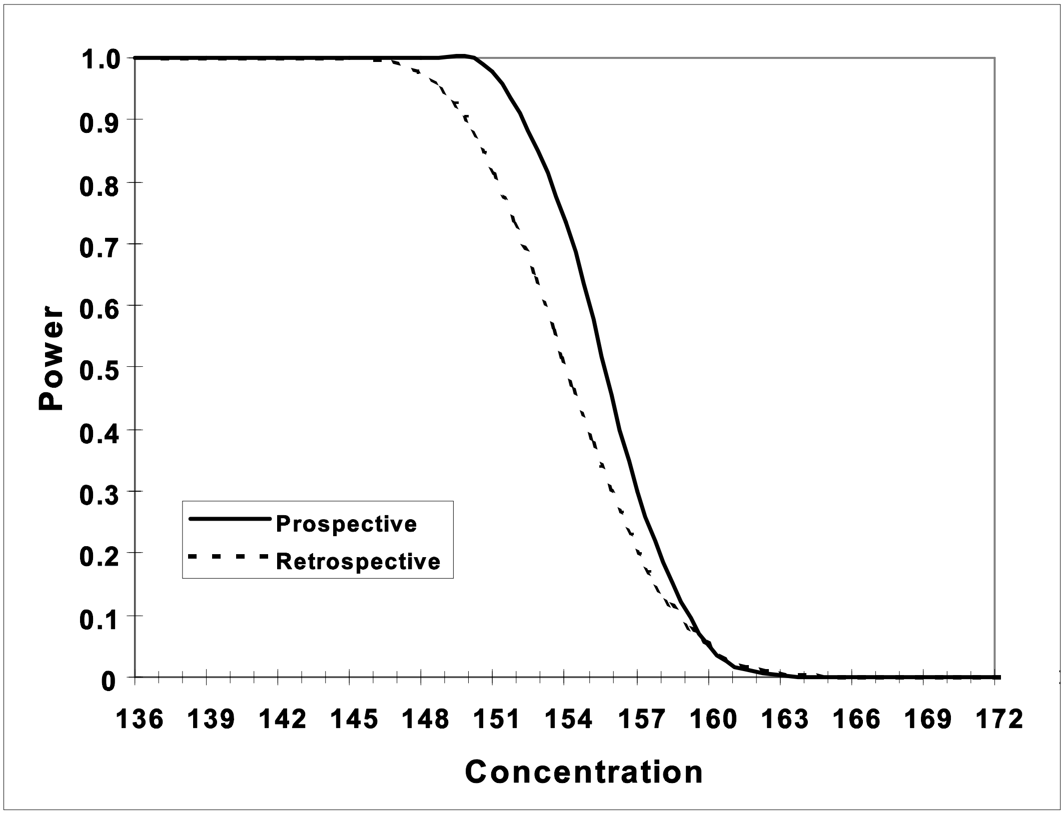

Figure 10.4 Example Power Curves: WRS Test Scenario A

10.5 Power of the Wilcoxon Rank Sum Test Under Scenario B

Recall that for the WRS test in Scenario B, the test statistic, W_s , was equal to the sum of the ranks of the survey unit measurements adjusted for the LBGR. If W_s exceeds the critical value W_c, then the null hypothesis that the median concentration in the survey unit exceeds that in the reference area by less than the LBGR is rejected, i.e., the survey unit does not pass this test.

The power of the WRS test in Scenario B can be approximated in a manner similar to that used in Scenario A, using Equations 10-5, 10-6 and 10-9:

$$\text{Power} = 1 - \Phi\left[\frac{W_c - 0.5 - 0.5m(m+1) - \mathrm{E}(W_{MW})}{\sqrt{\mathrm{Var}(W_{MW})}}\right]$$

$$\mathrm{E}(W_{MW}) = mnp_1$$

$$\mathrm{Var}(W_{MW}) = mnp_1(1-p_1) + mn(n+m-2)(p_2-p_1^2)$$

W_c is the critical value found in Table A.4 for the appropriate number of survey unit measurements, *n*, and number of reference area measurements, *m*. Since under Scenario B, both the WRS test and the Quantile test are used in tandem, the value of the Type I error, α, decided on during the DQO process, is halved for each test. Thus, the Table A.4 value for value of W_c for $\alpha_W = \alpha/2$ is used. $E(W_{MW})$ and $Var(W_{MW})$ are the mean and variance of the Mann-Whitney form of the WRS test statistic for Scenario B, namely $W_{MW} = W_s - n(n+1)/2$. Values of $\Phi(z)$, the standard normal cumulative distribution function, are given in Table A.1.

In Scenario B, p_1 is the probability that any single measurement from the survey unit exceeds a single measurement from the reference area by more than the LBGR . This probability depends on the difference in median concentration between the survey unit and the reference area. When this difference is equal to the $\mathrm{DCGL_W}$, then p_1 is equal to P_r as calculated from Equation 9-9. For other values of the difference median concentration between the survey unit and the reference area, C, we simply replace the $\mathrm{DCGL_W}$ in Equation 9-9 with C:

$$
\begin{aligned}
P_r(C) &= \text{Probability}(U = X - Y > LBGR) \\
&= \int_{LBGR}^{\infty}\left[\int_{-\infty}^{\infty} f_X(u+y) f_Y(y)\,dy\right] du \\
&= \int_{LBGR}^{-\infty}\left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(u+y-C-BKGD)^2/2\sigma^2} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-BKGD)^2/2\sigma^2} dy\right] du \\
&= \frac{1}{\sqrt{2\pi}\,\sqrt{2}\,\sigma} \int_{LBGR}^{\infty} e^{-(u-C)^2/4\sigma^2} du \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{C-LBGR}{\sqrt{2}\,\sigma}} e^{-x^2/2} dx \\
&= \Phi\left(\frac{C-LBGR}{\sqrt{2}\,\sigma}\right)
\end{aligned}
\qquad (10\text{-}10)
$$

This is the same as Equation 10-7, with $(DCGL_W - C)$ replaced by $(C - LBGR)$. Although the definition of p_1 has changed, its value may still be found from Table 10.3 when $(C - LBGR)/\sigma$ is substituted for $(DCGL_W - C)/\sigma$. Note that if $C = LBGR$ then $p_1(C) = 0.5$. The assumption of normality is not critical in the above calculations, since it is only being used to estimate the power. However, if a different distribution is considered more appropriate, Equation 9-10 can be used to calculate $p_1(C)$.

POWER

In Scenario B, p_2 is the probability that two random measurements from the survey unit will each exceed a single random measurement from the reference area by more than the LBGR; and p_3 is the probability that a single random measurement from the survey unit will exceed each of two random measurements from the reference area unit by more than the LBGR. When the difference in the concentration distributions of the survey unit and the reference area measurements consists of a shift in the median, and the measurement distributions are symmetric, then $p_2 = p_3$. If the measurement distributions are normal, then p_2 is equal to the probability that two correlated standard normal random variables (i.e., with mean = 0 and variance = 1), with correlation coefficient 0.5, are both less than $(C - LBGR)/(\sigma\sqrt{2})$. This probability also depends on the difference in median concentration, C, between the survey unit and the reference area. Again, values of p_2 may be obtained from Table 10.3 when (C−LBGR)/σ is substituted for $(DCGL_W - C)/\sigma$.

Although the power calculated as above is an approximation, this approximation has been compared against the power simulations for the WRS test reported by Gilbert and Simpson (PNL-7409, 1992). It was found that the approximation is sufficiently accurate to determine if the sample design achieves the DQOs.

When the values of $p_1(C)$ and $p_2(C)$ from Table 10.3, and the resulting $E(W_{MW})$ and $Var(W_{MW})$ are inserted in Equation 10-5, we obtain the probability that the null hypothesis is rejected at the concentration C. When $C = DCGL_W$, this probability is the probability of a Type I error, $\alpha_W = \alpha/2$.[5]. This calculation can even be performed for values of C less than the LBGR . The probability obtained is still the probability that the null hypothesis is rejected, i.e., that the survey unit passes the test, but it is not usually referred to as the power.

If the probability that the null hypothesis is rejected (calculated from Equation 10-5) is plotted against the concentration, C, the result is called a power curve. When the power calculation is performed at the design stage, using an estimated value of σ, it is called a prospective power curve. When the calculation is performed after the survey, using the standard deviation of the survey unit measurements as an estimate of σ, it is called a retrospective power curve.

To illustrate the construction of a power curve, consider the example of Chapter 6. The $DCGL_W$ for this example was 160 and the LBGR was 142. The DQOs for $\alpha_W = \alpha/2 = 0.025$, and $\beta = 0.05$, result in a sample size of $n = m = 12$, using the estimate that $\sigma = 6$. From Table A.4, the critical value for the WRS test with $n = m = 12$ and $\alpha = 0.025$ is $W_c = 184$. This is all of the information necessary to construct the prospective power curve. To construct the retrospective power curve, we use the larger of the standard deviations of the measurement data from the survey unit and the reference area, 8.1, as the estimate of σ.

The results of these calculations are shown in Table 10.5 and Figure 10.5.

---

[5] The value of α actually obtained from Equation 10-5 should be close to that specified in the DQOs. It may not exactly equal that value when the sample sizes are small, since the critical value, *k*, can only take integer values.

Table 10.5 Example Prospective Power Calculation: WRS Test Scenario B

C	$(C-LBGR)/\sigma$	p_1	p_2	$E(W_{MW})$	$Var(W_{MW})$	$SD(W_{MW})$	z	Power
136	−1.0	0.239750	0.113202	34.5	202.8	14.2	4.98	0.00
139	−0.5	0.361837	0.206266	52.1	271.9	16.5	3.24	0.00
142	0.0	0.500000	0.333333	72.0	300.0	17.3	1.93	0.03
145	0.5	0.638163	0.482593	91.9	271.9	16.5	0.82	0.20
148	1.0	0.760250	0.633702	109.5	202.8	14.2	−0.28	0.61
151	1.5	0.855578	0.765812	123.2	124.9	11.2	−1.58	0.94
154	2.0	0.921350	0.865767	132.7	63.9	8.0	−3.40	1.00
157	2.5	0.961450	0.931365	138.4	27.4	5.2	−6.29	1.00
160	3.0	0.983053	0.968795	141.6	10.0	3.2	−11.40	1.00
163	3.5	0.993336	0.987410	143.0	3.2	1.8	−21.15	1.00
166	4.0	0.997661	0.995497	143.7	0.9	0.9	−40.85	1.00
172	5.0	0.999796	0.999599	144.0	0.0	0.2	−175.07	1.00

In the figure it can be seen that the retrospective power is slightly less than that specified in the DQOs. However, in this example the null hypothesis was rejected, so the question of the power is moot. The retrospective power calculation is really only necessary when the null hypothesis is accepted. In that case it is important to know that it was not accepted simply because there was insufficient power. When the null hypothesis is rejected in spite of insufficient power, the survey designer can consider himself lucky, but the conclusion is still statistically valid.

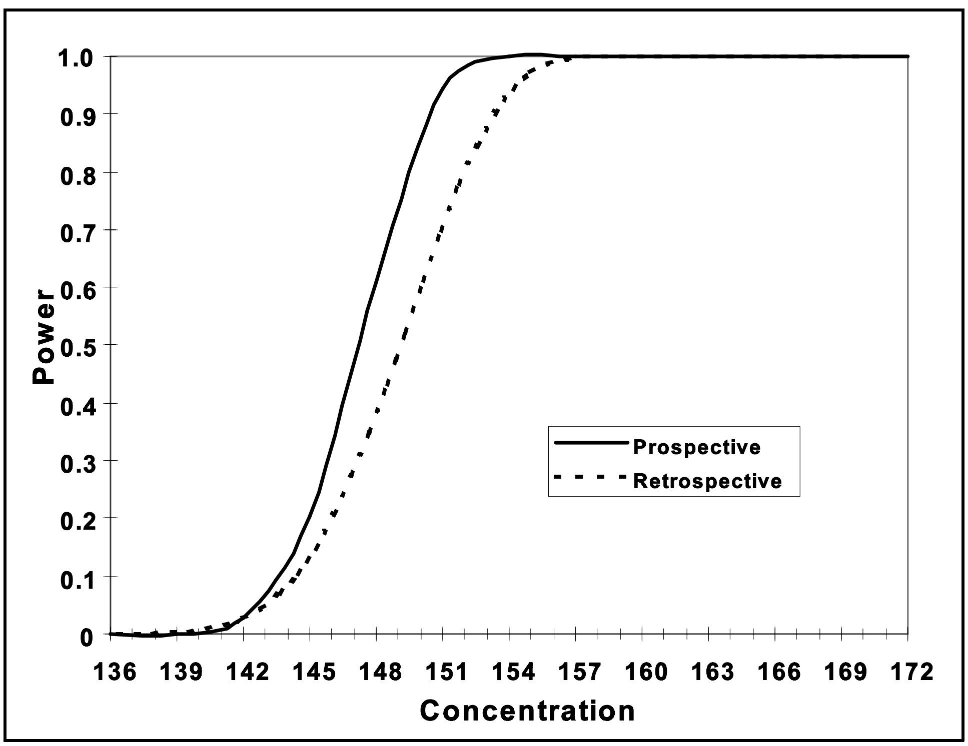

Figure 10.5 Example Power Curves: WRS Test Scenario B

In Scenario A, the power and the probability that the survey unit passes the test are equivalent. In

Scenario B, the power is equivalent to the probability that the survey unit does not pass. Thus, the probability that the survey unit passes is one minus the power. The result is plotted in Figure 10.6.

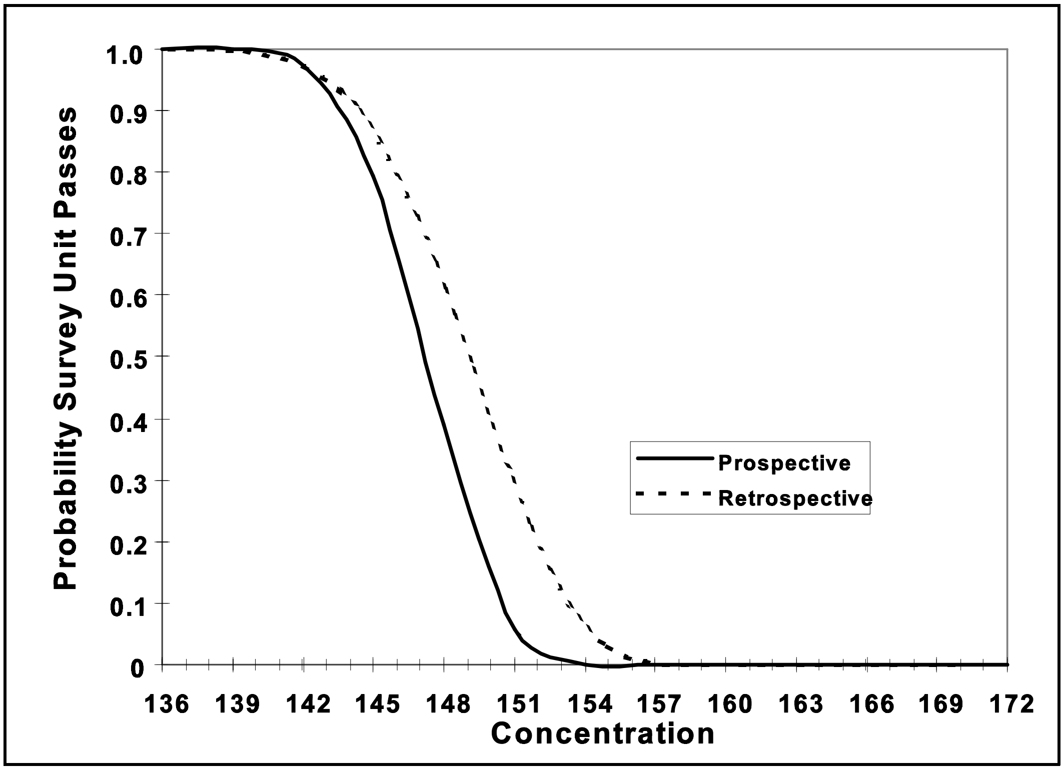

Figure 10.6 Probability Example Survey Unit Passes: WRS Test Scenario B